Mawuli K.E. Sevor

March, 2014

WHAT IS RESEARCH?

Research is a systematic process of data and information acquisition and analysis in order to increase understanding about a phenomenon we are interested in or concerned about.

|  |  |  |
| --- | --- | --- |
| **Steps** | **Activity** | **Method** |
| 1 | Hypothesis formulation | A statement about the relationship between two or more theoretical entities |
| 2 | Testing the hypothesis | Design a study or framework to examine whether the relationships between the entities are as hypothesised |
| 3 | Data collection | Process of collecting and preparing data pertaining to the study |
| 4 | Accept / Reject | Decide to **accept** or **reject** the hypothesis. |

Table 1: The Scientific Method summarised from John Dewey (1933)

The purpose of using a scientific method is to enable researchers describe the relations between factors; to predict what might happen; to control variables (when variables are manipulated, does it lead to certain cause or condition?) and to explain the investigated phenomena.

RESEARCH IS NOT

1. Merely information gathering. Example: a student goes to the library to do research and learned a lot about rice hybridisation. This student might think that research means going to the library to get information or to glance at a few facts. This may be information discovery; good for learning reference skills but it is not research.
2. Just transportation of facts/information from one location to another. This is more realistically called fact discovery or fact transportation. It misses the essence of research and the interpretation of data.
3. Merely rummaging for information from one’s personal records or at the library, is not research.

It is an exercise in self-enlightenment.

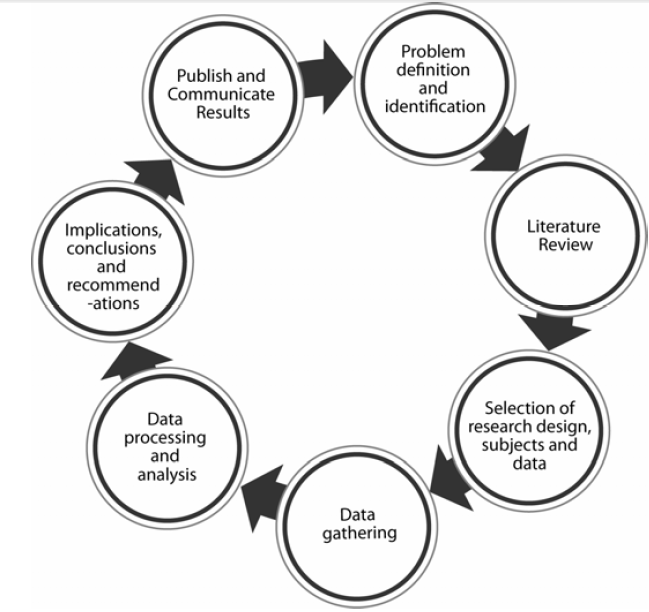
1. A catchword used to get attention: Research is not branding or “catchword” to get public attention for products or services. Example: A company using the phrase “years of research has produced this high quality product” is simply demonstrating the clever use of a catchword.

CHARACTERISTICS OF RESEARCH

* It starts with a question or problem in the mind of the researcher.
* It requires a clear goal.
* It requires a specific plan or procedure.
* It usually divides the principal problem into more manageable sub-problems.
* It is guided by the specific research problem, questions or hypothesis.
* It accepts certain critical assumptions.
* It requires the collection and interpretation of data in an attempt to resolve the problem that initiated the research.
* It is cyclical or more exactly helical in nature.

1. First of all, identify a problem in the field that is close to your interest or maybe closely related to your specialisation.
2. Review literature in that area of study. It can reveal similar investigations, show and suggest approaches in dealing with similar problems.

For example, if you are interested in farm waste management, you should start with textbooks in that area or chapters of books related to challenges of farm waste management. Textbooks could provide fundamental knowledge as a start in a particular area. However, journals and thesis could provide critical information as they review the related literature. Journals and thesis can reveal source of data which are never known to you. It could also introduce you to significant research personalities whose research and writings are not known to you. Generally, students tend to refer to other popular sources such as magazines and newspapers compared to journals and theses. Journals are presumed to be the original source of information into new technologies.

1. Research design is crucial as it provides overall structure for the research procedure, the data that a researcher collects and the data analysis that a researcher conducts. It is a plan that ensures the success of your research by identifying resources, procedures and data, from the very beginning.
2. Two types of data collection are available based upon the design that is suitable for your study. They are **quantitative** and **qualitative** data collection methods.
3. In any field of research, data collected need to be analysed and computed in order to provide us with the inferential and interpretation of the studied problem. If the research question involves reasoning, the quantitative approach, statistical methods are used in the analysis. The analysed data is usually presented in tables and graphs. A researcher then interprets the data in relation to the research questions based on the analysis performed. On the other hand, for qualitative approach, the information will be coded, justified and presented with valid reasoning.
4. Implications and conclusions are important justification that every researcher should take note as part of the research process. The novelty of the work and contribution of new knowledge will be seen in the “implications” and “conclusion” parts of the entire research. Recommendations normally highlight a few potential research questions which should be derived at the end of the research process and foster research continuation based on the new findings. That is why research is said to be cyclical, which means that research begets more research and one can come across additional problems that need resolving after a conclusion is reached.
5. Once the findings of the research are obtained, it is important for you to communicate with other fellow researchers and practitioners. The obtained results should be published in journals and conference proceedings in order that interested parties will communicate and discuss the findings. Student findings are communicated in the form of thesis / dissertation / long essay.

**Assignment:**

🡺What is engineering, exploratory, interpretivist, pure, instrumentalist, applied and problem-oriented research?

🡺Who is a rational positivist and who is a post-positivist?

ETHICS IN RESEARCH

Research is governed by a code of conduct or ethics that has evolved over the years to accommodate the values, needs and expectations of those who hold a stake in research.

**Ethical issues concerning research participants:**

🡺 **Informed Consent**

Respondents in research study should be told and informed about the nature of the study and given the choice of participating or not participating. A respondent also has the right to withdraw from the research study anytime they desired. This means that participation should be strictly voluntary.

In surveys, it is common practice to present an informed consent form that describes the nature of the research work as well as the nature of the respondent’s participation in it.

🡺 **Protection from Harm**

Researchers should not expose their respondents to physical or psychological harm. This applies in all situations especially during data collection period from respondents or during subjects’ involvement in experiment. A researcher should carefully examine whether respondents’ involvement is likely to cause harm. It is important to educate the respondents earlier before conducting any study or experiments.

🡺 **Confidentiality**

Sharing information about respondents in your research work for purposes other than research is unethical. Privacy and confidentiality are important. It is unethical to release your respondents’ information to third parties. Ensure that sources of information collected are protected. Maintain anonymity and confidentiality in treating respondents.

🡺 **Sensitive information permission**

Certain types of information can be regarded as sensitive and might not be convenient for a respondent in a research study. However, if your research work needs such crucial information to contribute new knowledge, seek respondents’ permission to obtain such information; give them sufficient time to decide on participation without any inducement.

**Ethical issues concerning the researcher:**

🡺 **Plagiarism**

The most important issue in academic and research integrity is plagiarism. Researchers who claim work done by someone else as their own work are very much unethical. Give proper acknowledgement for works that are not yours but cited in your work. Works include photographs, charts, graphs, drawings, statistics, interviews and texts published electronically or printed. To avoid plagiarism, paraphrase clauses, sentences or paragraphs in your own words and cite your source.

🡺 **Bias**

Bias is deliberate attempt either to hide what you have found in your studies or highlighting/emphasising something which is not true. Research should report and cite true findings honestly and without any bias.

METHODOLOGY OF RESEARCH

1. Design the question that you want to ask using observations of problems or challenges in the environment, reading published research papers, journals and looking out for recommendations or suggestions for further research.
2. Collect data or observations that help to answer the question which will require

* a research design 🡺 instrumentation 🡺 sampling strategy (2 groups = probability and non-probability).

Make sense of the data using statistical tools like graphs, Least Significant Difference (LSD), Duncan’s multiple range test, orthogonal comparisons, descriptive statistics, ANOVA, Chi Square, T-test, Correlation = Data analysis.

4. Use the results from data analysis to answer the research question = Interpretation of results.

The clearer sense we have of the research topic, the better chance we have of actually finding out what is of interest to us; hence the need for literature review at every stage of the research.

HYPOTHESIS TESTING

Assignment

1. What is a hypothesis?
2. What are Type I and II errors?

**Note:** All references must be properly cited and referenced according to APA style.

Type I errors (false positives leading to false scientific claims)?

Type II errors (false negatives resulting in missed scientific discoveries)?

Type II error rate is a function of several things including sample size (positively correlated with experiment cost), significance level (when the standard of proof is high, the chances of overlooking a discovery are also high) and effect size (when the effect is obvious to the casual observer, Type II error rates are low).

**Error:** Unexplained variation in a collection of observations

**Random error:** Error that occurs due to natural variation in the process. Also called experimental error, it is typically assumed to be normally distributed with zero mean and a constant variance.

**Experimental unit:** The entity to which a specific treatment combination is applied.

**Replication:** Performing the same treatment combination more than once in an experiment.

**Population**: the entire set of things (crops, animals, people, events) we want our research answer to apply to. Define who and or what falls within.

**Sample**: a subset of the population. Why?

**Statistic**: a numerical fact about a sample.

**Parameter:** a numerical fact about a population.

**Note:** For practical reasons, statistics are used to approximate population parameters in a well planned study. What are some of those practical reasons?

**Bias** occurs when the statistic we observe is not a good estimate of the population parameter.

**Observed score/value:** true value + chance error + bias

Assignment

1. Compare and contrast the law of averages, the law of large numbers and the law of small numbers.
2. Which of these will you recommend for research work?

REASONS FOR BIAS OCCURRENCE

**Poor sampling plans/strategy:**

Volunteer bias: volunteers for research studies tend to be better educated than those who do not.

Selection bias: we may select some elements of our population more easily than others.

Non-response bias: we may fail to select some elements at all from our population. For example, in a study about OAC students, those who have travelled at the time of data collection will not be sampled.

**Instrumentation problems:**

Response bias: people do not always answer questions accurately. Why?

Instrumentation bias: Badly calibrated measuring tools or faulty.

**Bias can occur because of design problems.** Give examples.

TYPES OF VARIABLES

Four types of scales: Nominal, Ordinal, Interval and Ratio. Nominal variables can only be counted so no mean or standard deviation can be calculated on such. Ordinal variables can be counted and median can be calculated. Interval and ratio (combined as scale in SPSS) are quantitative variables; most quantitative analysis can be performed on these. Qualitative variables do not have fixed distance but quantitative variables have fixed distance.

**Nominal:** Qualitative variable without order (only categorisation possible).

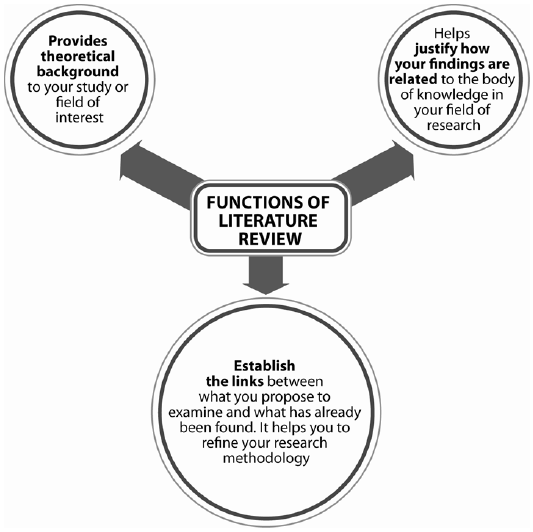
**Ordinal:** Qualitative variable with order/ranking (categorisation and order).

**Interval:** Quantitative variable without fixed origin but has fixed distance. Example: temperature.

**Ratio:** Quantitative variable with fixed origin and distance.

LITERATURE REVIEW

Literature review uses reports of primary or original scholarship, and does not report new primary scholarship itself. The primary reports used in literature may be verbal, but in the vast majority of cases reports are written documents. The types of scholarship may be empirical, theoretical, critical/analytic or methodological in nature. Secondly, literature review seeks to describe, summarise, evaluate, clarify and/or integrate the content of primary reports (Cooper, 1989).



Many students think that literature review means collecting and compiling facts for the research being undertaken. In fact, the literature review process needs analytical thinking, critiquing ability and empirical approach. The review of literature is a standard topic of a research report. Review of literature is an integral part of an entire research process. When you undertake a research process, review of literature will help you to establish the theoretical roots of your field of interest, clarify your ideas and develop your methodology. The review of literature also helps you to integrate your findings with the existing body of knowledge. Remember that one of your important responsibilities in research is to compare your findings with those of others, and that is why review of literature plays a very important role in the research process.

The aim of literature review is to highlight what has been done so far in the field of interest and how your findings relate to earlier research. The review of literature also indicates the approaches, methods, variables and statistical procedure used in a study. The review gives an overview on findings based on previous research work done. The review also traces the general patterns of the findings and the conclusions that can be made based on the findings. Generally, review of literature provides in-depth understanding and explanation on how your findings are similar to or novel from previous research work.

For example, literature review could justify whether your work is an extension of what others have done. It could also indicate whether you are trying to replicate earlier studies in a different context.

Review of literature also reveals techniques and statistical procedures that have not been attempted by others. To accomplish review of literature, you need to locate, read and evaluate research documents, reports as well as thesis and other sources of academic materials. Review for any research process must be extensive and thorough to present a detailed account of the topic being studied.

IMPORTANCE OF LITERATURE REVIEW

🡺 **Improve your Research Methodology:** Literature review helps you to acquire methodologies used by other researchers to find and solve research questions similar to the ones you are investigating. It explains the procedures other researchers used and methods similar to the ones you are proposing. It gives an idea of whether the methods other researchers used worked for them and the problems they faced. Literature review makes you aware of pitfalls and problems in order to strategise properly to select a methodology that better fits the research work.

🡺 **Focus on Research Problem:** Review of literature can help you shape your research problem because the process of reviewing the literature helps to understand the subject area better and thus helps you to conceptualise your research problem clearly and precisely. It also helps to understand the relationship between your research problem and body of knowledge in your research area.

🡺 **Cater to Knowledge Base for Research Area:** One of the most important objectives of literature review is to ensure that you read widely around the subject area in which you intend to conduct research. It is fundamental that you know what others are doing in your field of interest or topic you are working on as well as understanding theories that have been put forward and gaps that exist in the particular field.

🡺 **Contextualising Research Findings:** Obtaining answers for the research questions is easy. The difficulties lie in how to examine research findings in the existing body of knowledge. How do you answer your research questions compared to what other researchers concluded? What is the new knowledge contribution from your research work? How are your findings distinguished from those of other researchers? Answers to these questions require review of literature. It is important to put your findings in the context of what is already known and understood in your field of study.

🡺 **Ensure Novelty in your Work:** review of literature reduces the risk of wasting efforts on trying to rediscover what is already known or published in the research arena. Therefore, through literature review, you could ensure novelty and new contribution in your research work.

HOW TO REVIEW LITERATURE

🡺 **Step 1: Search the Existing Literature in your Research Area of Interest:** Make sure your chosen research area is well-researched and well studied which could give you more lines of research to choose from. Narrow your topics so that you can cover in depth. Comprehensive and narrowness of topic go hand in hand. Then, proceed to search the existing literature. To effectively search literature, have in mind some idea of the broad subject area and the problem you wish to investigate. The first task would be compiling a bibliography in your research area. Books and journals are always the best sources for literature in any research area. Sources can be: indices of journals, abstracts of articles, and citation indices. Modern libraries store information on printed and electronic media.

🡺 **Step 2: Review the Literature Obtained:** reading identified publications critically to pull together themes and issues that are associated with your research topic. Read and read! That is the bottom line.   
If you do not have a framework or theme to begin your research with, use a separate paper to jot down the main points you extract from journal articles and books. Once you create a rough framework, you may slot in the extracted information accordingly. As you read further, do some critical review with particular references to the following aspects:

1. Theories put forward, critics, methods used (sample size, data used, and measurement procedure).
2. Note whether the knowledge relevant to the designed framework is confirmed beyond doubt.
3. Find differences of opinions among researchers and jot down your opinions about their validity.
4. Examine the gaps that exist in the body of knowledge.

🡺 **Step 3: Develop a Theoretical Framework**

Research reports are limited by time, so set the boundary and parameters by looking into literature relevant to your topic. Information obtained from literature sources must be sorted according to themes and issues in your framework. Review literature with regard to the framework you developed in order to focus in your literature search. The theoretical framework provides a base and guide to read further; therefore, develop a framework first and then dive into literature search or vice versa. As you read more about your research area, you are likely to change the framework; it is a normal part of research.

🡺 **Step 4: Writing up the Literature Review**

Then compile and write all the literature you read and reviewed. Begin your review with a theme or points that you want to emphasise. Organise and list all the themes you would like to discuss and relate. Organisation is very important and makes the structure known to your reader. While writing, identify and describe various theories relevant to your field and specify gaps in the knowledge of body in that area. Proceed to explain recent advances in the area of study as well as current trends. In agriculture research, current trends are always given great importance. Write the review using descriptions, comparisons and findings evaluation based on:

(a) Assumptions of research (b) Theories and history related to the area of study

(c) Hypotheses (d) Research designs applied

(e) Variables selected (f) Potential future work speculated by researchers

Most importantly, avoid plagiarism when writing. Give due recognition to the works of other researchers. Quote from other researchers’ work to show how your findings contradict, confirm or add to theirs after finalising your data analysis during the research process. Acknowledging sources shows the precision, breadth and depth of your review.

COMMON MISTAKES IN LITERATURE REVIEW

Most beginners in research make the following mistakes as soon as they start reviewing literature:

1. The review made is a mere description of various materials without making an effort to show the relation between the studies and main objective of the research topic.
2. Cut and paste, which SHOULD NOT be encouraged. Original works should be cited and quoted.
3. Journals or reports that are included are not critically evaluated. Critically evaluate the research questions, the methodology used and recommendations made by the researcher.

**Note:** It has been realized that students have not read the original works but instead cite a secondary source as though they had read the primary source.

**Assignment:** Compile and review literature on the topic: Literature review.

RESEARCH DESIGN

Research design is a procedural plan that is adopted by a researcher to answer questions in a valid way. It is very objective, accurate and defines what can be done or not. Normally, a research design determines the type of analysis to do to get the desired results. To what extent your design is good or bad depends on whether you are able to get the answers to your research questions. If your design is poor, the results of the research will also be bad.

How do you go about getting a good research design that will provide sufficient answers to your research work? There is no fixed method to guide us on how to do it. The best approach will be to examine different types of research designs, list down the strengths and weaknesses and then make your decisions.

Researchers, must have good understanding of the research/problem, and be clear about research questions and what it is to establish. It is good practice to understand the challenges in every research before going in depth. Never select a design and then try to fit the research questions to the design! It should be vice versa. It is very important to assess whether the design can answer the research questions. So, choose a design that will give optimised results.

A research design relates to the identification of procedures and logistical arrangements to start a study and also at the same time emphasises the importance of quality in producing optimised research results; it glues all the components and subcomponents in a research project together. A research design helps to measure quality and reliability of the study.

Generally, the functions of research design are to:

* Construct an operational procedure to execute the tasks required in completing the research work.
* Ensure these procedures are sufficient to get valid and objective answers which are accurate to the questions posed in the research work.

A fundamental aspect of research design is to specify everything in depth and clear. This is to ensure that a reader will understand what method to follow and how to follow it. A good research design should:

* Name the study design (for example as comparative, cross-sectional or random control)
* Define and operationalise the key words in the research
* Identify the study population
* Define the sampling methods/procedure
* What method of data collection will be used in the research work?
* How will ethical issues be considered?

**Assignment:** Explain “after-only design”, “before-and-after design”, “post-test” and “pre-test”.

MEASUREMENT, OPERATIONALISATION & CONCEPTUALISATION

Research begins with concepts and then translates these into specific elements for data collection and analysis. Any form of research involves a process in which a researcher takes a general concept or idea, specifies the dimensions that he/she wants to study, then creates measures to evaluate these dimensions. These are the processes known as conceptualization and operationalisation.

In evaluating information in any topic (eg., topic X), it is helpful to know: "What is meant by X in this research?" and "How was X measured?" Researchers must develop answers to both questions before they proceed with their research.

Conceptualization is the process of specifying what is meant by a term. It is done by taking portions of an abstract theory and translating them into testable hypotheses involving specific elements. Think about the case of poverty. Every poverty research must create a specific conceptualization of the term poverty.

Criterion-Related Measure

There are a number of different ways to create specific, testable measures of a concept/phenomenon. With poverty as an example: We could employ a criterion-related measure where the operationalisation is examined to find out whether it behaves the way it should, given the theory of the construct. It assumes that the operationalisation will function in predictable ways in relation to other operationalisations based on your theory of the construct. Perhaps the most common measure of poverty involves comparing family income against the "poverty line". This poverty measure was developed in the mid-1960s and is defined in terms of the amount of money required to purchase an emergency amount of food a family needs and other essential goods.

Alternative Criterion

The alternative criterion approach does not compare income against a scale, it rather asks about other criteria. An increasing number of studies use criterion measures related to elements of material deprivation. Since poverty reduces families' ability to pay for essential goods, we can get a measure of poverty by asking how often families had to do without certain things, such as food or clothing.

Related Measure

We could also create a relative measure of poverty; for example, defining the lowest portion of an income distribution as living in poverty.

Subjective Approach

For example, asking individuals about how poor they feel. For some questions, it may be more important to know about whether people feel impoverished, relative to their peers, than to know about how their income compares to a threshold. Using these specific conceptualized constructs and translating them into specific measures (such as interview questions) that can be used to collect data is operationalisation.

From these definitions, we can think about a two-stage process of social research: we begin with a concept, develop specific dimensions of interest through conceptualization, and then create questions or specific measures through operationalisation. Listed below are two examples of the process for poverty:

|  |  |  |
| --- | --- | --- |
| Concept | Dimension of Interest | Specific Questions or Measures |
| Poverty | Absolute or "Threshold" Poverty | "How much money did your family earn last year?" (followed by a comparison of this income to a threshold) |
| Poverty | Subjective Poverty | "How poor would you say you feel?" |

There are many other different ways to ask about poverty besides these as well. We could ask individuals whether their family received food cards or whether their children received free lunch at school (since food cards and free lunch programs are only available to families who live below the poverty line). We could ask how often they went hungry or without another necessity (since, presumably, only those who could not afford such things would do without). Decisions about which of these measures are best for a particular research project or study are driven largely by the specific elements of most interest to researchers. Therefore the purposes of the study itself shape decisions about how to conceptualize and operationalise any particular topic.

VALIDITY AND RELIABILITY

Two of the primary criteria of evaluation in any measurement or observation are:

* Whether we are measuring what we intend to measure = Validity
* Whether the same measurement process yields the same results = Reliability

Validity refers to the extent to which we are measuring what we want to measure. Are we using the right measurement tool, is our tool good?

Reliability is concerned with questions of stability and consistency; ie: does the same measurement tool yield stable and consistent results when repeated over time? Think about measurement processes in other contexts - in construction or woodworking, a tape measure is a highly reliable measuring instrument.

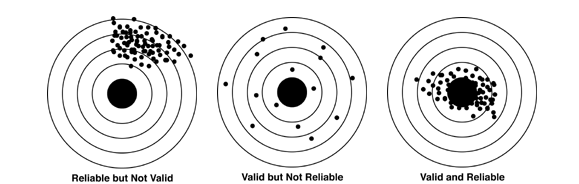
To apply these concepts to social research, we want to use measurement tools that are both reliable and valid. We want questions that yield consistent responses when asked multiple times - this is reliability. Similarly, we want questions that get accurate responses from respondents - this is validity.

Reliability

Reliability refers to a condition where a measurement process yields consistent scores (given an unchanged measured phenomenon) over repeat measurements. Perhaps the most straightforward way to assess reliability is to ensure that they meet the following three criteria of reliability. Measures that are high in reliability should meet all three.

* Test-Retest Reliability

When a researcher administers the same measurement tool multiple times (asks the same question, follows the same research procedures, etc.) does he/she obtain consistent results, assuming that there has been no change in whatever he/she is measuring? This is really the simplest method for assessing reliability - when a researcher asks the same person the same question twice ("What's your name?"), does he/she get back the same results both times. If so, the measure has test-retest reliability. Measurement of the piece of wood talked about earlier has high test-retest reliability.

* Inter-Item Reliability

This is a dimension that applies to cases where multiple items are used to measure a single concept. In such cases, answers to a set of questions designed to measure a single concept (e.g., hunger) should yield equivalent answers.

* Inter-observer Reliability

Inter-observer reliability concerns the extent to which different interviewers or observers using the same measure get equivalent results. If different observers/interviewers use the same instrument to score the same thing, their scores should match. For example, the inter-observer reliability of a student presentation assessment is often checked by letting two observers use the same assessment tool to score student performance. If the instrument has high inter-observer reliability, the scores of the two evaluators should match.

Validity

Validity refers to the extent we are measuring what we hope to measure (and what we think we are measuring). To assess the validity of a set of measurements, a valid measure should satisfy four criteria.

* Face Validity

An assessment of whether a measure, on the face of it, measures the concept it is intended to measure. This is a very minimum assessment - if a measure cannot satisfy this criterion, then the other criteria are inconsequential. We can think about observational measures of behaviour that would have face validity. For example, striking out at another person would have face validity for an indicator of aggression. Similarly, offering assistance to a stranger would meet the criterion of face validity for helping. However, asking farmers about their favourite crop to measure poverty has little face validity.

* Content Validity

It is the extent to which a measure adequately represents all facets of a concept. Consider a series of questions that serve as indicators of depression (don't feel like eating, lost interest in things usually enjoyed, etc.). If there were other kinds of common behaviours that mark a person as depressed that were not included in the index, then the index would have low content validity since it did not adequately represent all facets of the concept.

* Criterion-Related Validity

Criterion-related validity applies to instruments that have been developed for use as indicator of specific trait or behaviour, either now or in the future. For example, think about the driving test as a social measurement that has pretty good predictive validity. That is to say, an individual's performance on a driving test correlates well with his/her driving ability.

* Construct Validity

But for many things we want to measure, there is not necessarily a pertinent criterion available. In this case, turn to construct validity, which concerns the extent to which a measure is related to other measures as specified by theory or previous research. Does a measure stack up with other variables the way we expect it to? A good example of this form of validity comes from early self-esteem studies.

Clinical observations in psychology had shown that people who had low self-esteem often had depression. Therefore, to establish the construct validity of the self-esteem measure, the researchers showed that those with higher scores on the self-esteem measure had lower depression scores, while those with low self-esteem had higher rates of depression.

Validity and Reliability Compared

So what is the relationship between validity and reliability? The two do not necessarily go hand-in-hand. At best, we should have a measure that has both high validity and high reliability. It yields consistent results in repeated application and it accurately reflects what we hope to represent. It is possible to have a measure that has high reliability but low validity - one that is consistent in getting bad information or consistent in missing the mark. It is also possible to have one that has low reliability and low validity - inconsistent and not on target. Finally, it is not possible to have a measure that has low reliability and high validity - you can't really get at what you want or what you're interested in if your measure fluctuates wildly.

SAMPLES AND SAMPLING

Introduction

"Statistical designs always involve compromises between the desirable and the possible."

Leslie Kish. *Statistical Designs for Research*. 1987. (New York: John Wiley and Sons) p. 1.

The quote above shows that all research designs involve some form of compromise or adjustment. One of the dimensions on which such compromises are made relates to the populations about which we wish to learn. There are many research questions we would like to answer that involve populations that are too large to consider learning about every member of the population. For example, how do Ghanaians feel about the job that the Ministry of Food and Agriculture is doing? Questions such as this are important in understanding the world around us, yet it would be impractical, if not impossible, to measure the feelings about MOFA of all Ghanaians. Generally, in answering questions, researchers examine a fraction of the possible population of interest, drawing statistical inferences from this fraction. The selection process used to draw such a fraction is known as sampling, while the group contained in the fraction is known as the sample.

Why do we Sample?

* Reduced cost: It is less costly to obtain data for a selected subset of a population, rather than the entire population. Furthermore, data collected through a carefully selected sample are highly accurate measures of the larger population. Public opinion researchers can usually draw accurate inferences for the entire population of Ghana from interviews of only 1,000 people.
* Speed: Observations are easier to collect and summarize with a sample than with a complete count. This consideration may be vital if the speed of the analysis (time frame/deadline) is important.
* Greater scope: Sometimes highly trained personnel or specialized equipment limited in availability must be used to obtain the data. A complete census (enumeration) is not practical or possible. Thus, surveys that rely on sampling have greater flexibility regarding the type of information that can be obtained. Keep in mind that the primary point of sampling is to create a small group from a population that is as similar to the larger population as possible. One of the features we look for in a sample is the degree of representativeness. Meaning, how well does the sample represent the larger population from which it was drawn? How closely do the features of the sample resemble those of the larger population?

Sampling Terminology

Samples are always drawn from a population. By "population" we denote the aggregate (the whole) from which the sample (the part) is drawn. The population to be sampled (the sampled population) should coincide with the population about which information is wanted (the target population). Sometimes, for reasons of practicality or convenience, the sampled population is more restricted than the target population. In such cases, precautions must be taken to ensure that the conclusions only refer to the sampled population.

Before selecting the sample, the population must be divided into parts called sampling units. These units must cover the whole of the population and they must not overlap, in the sense that every element in the population belongs to one and only one unit. In sampling individuals in a town, the unit might be an individual person, the members of a family, or all persons living in the same city block. In sampling an agricultural crop, the unit might be a field, a farm, or an area of land whose shape and dimensions are at our disposal. The construction of this list of sampling units, called a frame, is often one of the major practical problems. A randomly selected, small percent of a population of people can represent the attitudes, opinions, or projected behaviour of all of the people, if the sample is selected correctly.

Types of Sampling

Two categories: probability samples and non-probability samples.

* Probability Samples

The idea behind this type is random selection. More specifically, each sample from the population of interest has a known probability of selection under a given sampling scheme. There are four categories of probability samples described below.

* + - Simple Random Sampling

The most widely known type of a random sample is the Simple Random Sampling (SRS) using a probability of selection that is the same for every unit in the population. Simple random sampling is a method of selecting n units from a population of size N such that every member has equal chance of being selected.

Imagine a survey of 100 farmers in a small town with a population of 1,000 eligible farmers. We could write the names of all farmers on pieces of paper, put all pieces of paper into a box and draw 100 tickets at random. You shake the box, draw a piece of paper and set it aside, shake again, draw another, set it aside, until we have 100 slips of paper. Those selected = our sample. This sampling procedure is Simple Random Sampling; at each draw, every name in the box had the same probability of being chosen.

Note however that it will be difficult if not impossible in most cases to list everything on pieces of paper, put in a box, and then randomly draw numbers until desired sample size is reached. Although simple random sampling is ideal for social research and most of the statistics used are based on assumptions of SRS, in practice, SRSs are rarely seen. It can be terribly inefficient, and particularly difficult when large samples are needed. Other probability methods are more common.

* + - Stratified Random Sampling

In this form of sampling, the population is first divided into two or more mutually exclusive segments based on some categories of variables of interest in the research. It is designed to organize the population into homogenous subsets before sampling, then drawing a random sample within each subset. With stratified random sampling the population of N units is divided into subpopulations of units respectively. These subpopulations, called strata, are non-overlapping and together they comprise the whole of the population. When these have been determined, a sample is drawn from each, with a separate draw for each of the different strata. The sample sizes within the strata are denoted by respectively. If a SRS is taken within each stratum, then the whole sampling procedure is described as stratified random sampling.

The primary benefit of this method is to ensure that cases from smaller strata of the population are included in sufficient numbers to allow comparison. Say that you're interested in how job satisfaction varies by crop produced among a group of farmers in a town. To explore this issue, we need to create a sample of all farmers in the town. However, the farmer population in this town is predominantly cassava farmers, as the chart illustrates. If we were to take a simple random sample, there's a good chance that we would end up with very small numbers of Maize, Cowpea and Cocoa farmers. That could be disastrous for our research, since we might end up with too few cases for comparison in one or more of the smaller groups. Rather than taking a simple random sample from the population, in a stratified sampling design, we ensure that appropriate numbers of elements are drawn from each farmer group in proportion to the percentage of the population as a whole. Say we want a sample of 1000 farmers - we would stratify the sample by group then randomly draw out 750 farmers from the Cassava, 90 from the Maize, 100 from the Cowpea, and 60 from the Cocoa groups. This yields a sample that is proportionately representative of all the farmers as a whole.

**Why do we stratify?**

1. If data of known precision are wanted for certain subpopulations, then each of these should be treated as a population in its own right.
2. Administrative convenience may dictate the use of stratification; for example, an agency administering a survey has regional offices, which can supervise the survey for a part of the population (region).
3. Sampling problems may be inherent with certain sub populations, such as people living in institutions (e.g. hotels, hospitals, prisons).

Stratification may improve the estimates of characteristics of the whole population. It may be possible to divide a heterogeneous population into sub-populations, each of which is internally homogenous. If these strata are homogenous, i.e., the measurements vary little from one unit to another; a precise estimate of any stratum mean can be obtained from a small sample in that stratum. The estimate can then be combined into a precise estimate for the whole population.

There is also a statistical advantage in the method, as a stratified random sample nearly always results in a smaller variance for the estimated mean or other population parameters of interest.

* + - Systematic Sampling

It sounds very different from SRS but a variant of SRS that involves some listing of elements and drawing every nth element in the list as a sample. Example: a sample of 1,000 out of a list of 10,000 people.

Creating such a sample includes three steps:

1. Divide number of cases in the population by the desired sample size. Example: 10,000 ÷ 1,000 = 10.
2. Select a random number between 1 and the value from Step 1. Example: a number between 1 and 10 - say we pick 7.
3. Starting with number chosen in Step 2, take every tenth record (7, 17, 27, ...)

More generally, suppose that the N units in the population are ranked 1 to N in some order (e.g., alphabetic). To select a sample of n units, we take a unit at random, from the 1st k units and take every kth unit thereafter.

The advantages of systematic sampling method over simple random sampling include:

It is easier to draw a sample and often easier to execute without mistakes. This is a particular advantage when the drawing is done in the field.

Intuitively, it seems systematic sampling is more precise than SRS. In effect it stratifies the population into n strata, consisting of the 1st k units, the 2nd k units, and so on. Thus, we might expect the systematic sample to be as precise as a stratified random sample with one unit per stratum. The difference is that with the systematic one the units occur at the same relative position in the stratum whereas with the stratified, the position in the stratum is determined separately by randomization within each stratum.

* + - Randomized Composite Sampling (Planting materials)

Randomized composite sampling implies the division and classifications of a population into a number of units of the same make/nature. In the context of farm experiments, Randomized composite sampling stands for the breaking down of planting materials into a number of uniform, well- defined and homogeneous units to be distributed to the experimental plots. It ensures a fair and even distribution of planting materials, so that all the best or otherwise do not fall on one plot alone. The Randomized composite sampling is a method that ensures that the planting material to any plot is governed wholly by the laws of chance.

* Procedure:

1. Measure out enough seeds required for a block (about 20g). This = sample.
2. Mix thoroughly and divide into convenient number of heaps (about 5) called sub- samples.
3. Mix each heap and divide equally into as many sub-sub samples as the number of plots on the block.
4. Bulk all sub-sub samples bearing the same mark separately.
5. Allocate each bulk sample to a plot on that block. Repeat the process for the remaining blocks**.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **BLOCK 1** | Bulk all sub samples marked 1 | Bulk all sub samples marked 2 | Bulk all sub samples marked 3 | Bulk all sub samples marked 4 |
| **PLOT 1** | **PLOT 2** | **PLOT 3** | **PLOT 4** |

**NOTE:** Allocation of the numbers 1, 2, 3 and 4 to sub-sub samples should be randomized.

* + - Cluster Sampling

In some instances the sampling unit consists of a group or cluster of smaller units that we call elements or subunits (these are the units of analysis for your study). There are two main reasons for the widespread application of cluster sampling. The first intention may be to use the elements as sampling units, because there may be no reliable list available of elements in the population and it would be prohibitively expensive to construct such a list. In many countries there are no complete and updated lists of the people, the houses or the farms in any large geographical region.

Important things about cluster sampling:

1. Most large scale surveys are done using cluster sampling;
2. Clustering may be combined with stratification, typically by clustering within strata;
3. In general, for a given sample size, n cluster samples are less accurate than the other types of sampling in the sense that the parameters you estimate will have greater variability than an SRS, stratified random or systematic sample.

Non-probability Sampling

Social research is often conducted in situations where a researcher cannot select the kinds of probability samples used in large-scale social surveys. For example, you wanted to study hunger - there is no list of hungry individuals nor are you likely to create such a list. However, you need to get some kind of a sample of respondents in order to conduct your research. To gather such a sample, you would likely use some form of non-probability sampling. The primary difference between probability methods of sampling and non-probability methods is that in the latter you do not know the likelihood that any element of a population will be selected for study.

There are four primary types of non-probability sampling methods:

* Availability Sampling

A method of choosing subjects who are available or easy to find. It is sometimes referred to as haphazard, accidental, or convenience sampling. The primary advantage of the method is that it is very easy to carry out, relative to other methods. A researcher can stand at the hospital entrance or at the market and hand out surveys or wait for class to begin in a school and use all students present in the class. Advantages include – ease of use, particularly with a captive audience, and in some cases you can attain a large number of interviews through this method.

The primary problem with availability sampling is that you can never be certain what population the participants in the study represent. The population is unknown, the method for selecting cases is haphazard, and the cases studied probably don't represent any population you could come up with.

However, there are some situations in which this kind of design has advantages - for example, survey designers often want to have some people respond to their survey before it is given out in the "real" research setting as a way of making certain the questions make sense to respondents (called pre-testing). For this purpose, availability sampling is not a bad way to get a group to take a survey, though in this case researchers care less about the specific responses given than whether the instrument is confusing or makes people feel bad.

Despite the known flaws with this design, it is remarkably common but it is entirely unknown what population the results of such polls represent. At best, a researcher could make some conditional statement about people who are present in the market at a particular point in time who cared enough about the issue in question to wait and answer your questions.

* Quota Sampling

It is designed to overcome the most obvious flaw of availability sampling. Rather than taking just anyone, you set quotas to ensure that the sample you get represents certain characteristics in proportion to their prevalence in the population. Note that for this method, you have to know something about the characteristics of the population ahead of time. Say you want to make sure you have a sample proportional to the population in terms of gender - you have to know what percentage of the population is male and female, then collect sample that matches yours.

The primary problem with this form of sampling is that even when we know that a quota sample is representative of the particular characteristics for which quotas have been set, we have no way of knowing if sample is representative in terms of any other characteristics. If we set quotas for gender and age, we are likely to attain a sample with good representativeness on age and gender, but one that may not be very representative in terms of income and education or other factors.

Moreover, because researchers can set quotas for only a small fraction of the characteristics relevant to a study, quota sampling is really not much better than availability sampling. To reiterate, you must know the characteristics of the entire population to set quotas; otherwise there's not much point to setting up quotas. Finally, interviewers often introduce bias when allowed to self-select respondents, which is usually the case in this form of research. In choosing males 18-25, interviewers are more likely to choose those that are better-dressed, seem more approachable or less threatening. That may be understandable from a practical point of view, but it introduces bias into research findings.

* Purposive Sampling

Purposive sampling is a sampling method in which elements are chosen based on purpose of the study. Purposive sampling may involve studying the entire population of some limited group (Diploma students at OAC) or a subset of a population (graduates of OAC in 2010 who work with Cocobod). As with other non-probability sampling methods, purposive sampling does not produce a sample that is representative of a larger population, but it can be exactly what is needed in some cases - study of organization, community, or some other clearly defined and relatively limited group.

* Snowball Sampling

Snowball sampling is a method in which a researcher identifies one member of some population of interest, speaks to him/her, and then asks that person to identify others in the population that the researcher might speak to. This person is then asked to refer the researcher to yet another person, and so on. Snowball sampling is very good for cases where members of a special population are difficult to locate. For example, working with cases where it is difficult for people to open up.

The method also has an interesting application to group membership - if you want to look at pattern of recruitment to a community organization over time, you might begin by interviewing fairly recent recruits, asking them who introduced them to the group. Then interview the people named, asking them who recruited them to the group.

The method creates a sample with questionable representativeness. A researcher is not sure who is in the sample. In effect snowball sampling often leads the researcher into a realm he/she knows little about. It can be difficult to determine how a sample compares to a larger population. Also, there's an issue of who respondents refer you to - friends prefer referring to friends; they are less likely to refer to ones they don't like or fear.

HYPOTHESIS TESTING

About Hypothesis Testing

An essential component of the scientific process is the formulation and evaluation of hypotheses. In seeking to learn more about the social world, social scientists ask many different kinds of questions about relationships between factors of social life. How do investors change their behaviour when market conditions change? What role did political and social factors play in the 1983 hunger? Does the issue of government allowance (money) influence students' performance in school? To address these questions, social scientists form hypotheses which they then evaluate using some form of data.

The natural sciences evaluate hypotheses such as:

1. The combination of certain chemical compounds yields heat energy.
2. Plants' growth is enhanced through exposure to ultraviolet light.
3. When a moving object collides with another object, the total kinetic energy of the two objects does not change.

Typically, hypotheses such as these are generated from some theory or theoretical perspective, and then evaluated using data collected through laboratory procedures; research in the social sciences works similarly (though often outside the laboratory).

What is a Hypothesis?

A hypothesis is an empirically-testable statement about a relationship involving two or more variables. Examples of hypotheses from the social sciences include:

1. Graduates of Agricultural colleges suffer in getting employment.
2. Rural farmers are not consulted in making agricultural policies.
3. Students' allowance payment is an essential element of their academic success.

Each of these specifies a relationship that may or may not exist under particular conditions. They are testable statements about relationships between different factors. But why bother with forming a hypothesis as part of the research process?

Why Use Hypotheses in Research?

In examining phenomena of the social world, there are a number of relationships that can be examined to learn more. However, it is possible that some of the relationships that we observe might be due to chance, rather than some relationship between two variables. Establishing and testing hypotheses is a method used to evaluate the likelihood that relationships occurred by chance.

**Generating and Testing Hypotheses**

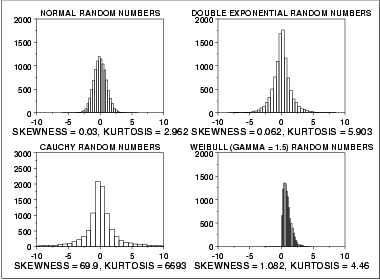
We establish two competing hypotheses that we then evaluate in light of some empirical data. These hypotheses are referred to as the null hypothesis and the alternative hypothesis. The primary purpose of hypothesis testing is to examine the likelihood of the null hypothesis with data.

The null hypothesis is often the reverse of what the experimenter actually believes; it is put forward to allow the data to contradict it. In a study of the effect of sense of control on health, the researchers expect that a sense of control will improve health. The null hypothesis they would establish in this setting, then, is that enhancing sense of control will have no effect on health. The alternative hypothesis is one that stands in contrast to the null, usually that the condition or change will have some effect. In the example mentioned above, the alternative hypothesis is that changes in sense of control will result in a change in health.

Depending on the data, the null hypothesis either will or will not be rejected as a viable possibility. If the data show a sufficiently large effect of the sense of control, then the null hypothesis that sense of control has no effect can be rejected. With this understanding of the way that hypotheses are generated in the social sciences, let us look at specific tools that are used to test hypotheses.

DATA ANALYSIS AND INTERPRETATION

SKEW & KURTOSIS

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the centre point.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set.

Data that has a **long tail** on one side or the other = Skewed data.

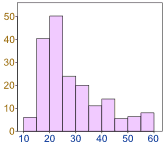
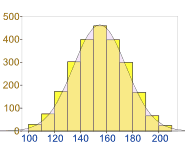
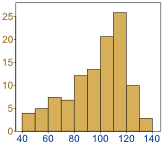
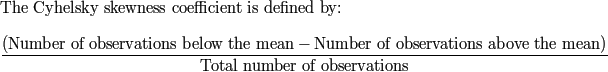


Figure 1: Negative Skew Figure 2: No Skew Figure 3: Positive Skew

### ANALYSIS 1

Plot a column chart using Data 3. Calculate the skew of the data to 1 decimal place using the SKEW() function. Is it positive or negative? Interpret the result of the analysis.

**Solution**

  
Calculate the Cyhelsky skewness coefficient for the set of numbers:  
11, 14, 17, 18, 27, 27, 29, 31, 38, 39

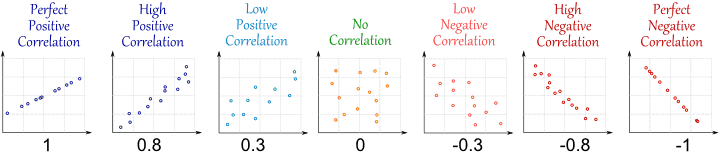
**Note:** Symmetric (normal) distribution: mean = median = mode. Positively skewed: mean > median. Negatively skewed distribution: mean < median.

Question: What is excess kurtosis?

CORRELATION

When two sets of data are strongly linked together we say they have a **High Correlation**.

* Correlation is **Positive** when the values **increase** together, and
* Correlation is **Negative** when one value **decreases** as the other increases



Correlation values can be positive or negative numbers between -1 and 1 or zero.

* 1 is a perfect positive correlation
* 0 is no correlation (the values don't seem linked at all)
* -1 is a perfect negative correlation

**Note:** The value shows **how good the correlation is** (not how steep the line is), and whether it is positive or negative.

### ANALYSIS 1

Plot a scatter chart using Data 1. Calculate the correlation of the data to 1 decimal place using Pearson’s Correlation. How good is the correlation? Is it positive or negative? Interpret the result of the analysis.

**Solution**

**Spreadsheet (OpenOffice Calc and Excel)** **CORREL() Function Calculation**

Formula looks like CORREL(range of data set 1, range of data set 2). Eg: **=CORREL (A3:A10, B3:B10)**

**Manual Calculation**

Let us call the two sets of data "x" and "y" (So, Temperature = **x** and Chilled Drinks Sales = **y**):

* Step 1: Find the mean of **x**, and the mean of **y**
* Step 2: Subtract the mean of x from every x value (label as "**a**"), do the same for y (label as "**b**")
* Step 3: Calculate: **a × b**, **a2** and **b2** for every value
* Step 4: Sum up **a × b**, sum up **a2** and sum up **b2**
* Step 5: Divide the sum of **a × b** by the square root of **[(sum of a2) × (sum of b2)]**

Can you put it all together in one formula?

### ANALYSIS 2

Plot a scatter chart using Data 2. Calculate the correlation of the data to 1 decimal place using Pearson’s Correlation. How good is the correlation? Is it positive or negative? Interpret the result of the analysis? Does correlation mean causation? Does correlation calculation work well on curves?

The Chi-Square Test

Introduction

One of the most common and useful ways to look at information is in the format of a table. For example, if we want to know whether men or women do more poultry production in Ohawu, we can use a table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Poultry** | **No poultry** | **Total** |
| **Women** | 46 | 71 | 117 |
| **Men** | 37 | 83 | 120 |
| **Total** | **83** | **154** | **237** |

The above example is relatively straightforward in that we can fairly quickly tell that more women than men kept poultry in Ohawu. Calculating percentages, we find that 39% of women kept poultry, as compared with 31% of men. However, to re-frame the issue, what if we wanted to test the hypothesis that women kept more poultry than men in Ohawu? These figures are a good start to examining that hypothesis; however, the figures in the table are only descriptive. To examine the hypothesis, we need to employ a statistical test, the chi-square test.

The Chi-Square Test

Generally speaking, the chi-square test is a statistical test used to examine differences with categorical variables. There are a number of features of society we characterize through categorical variables (variables that can take on one of a limited, and usually fixed, number of possible values) like religion and political preference. To examine hypotheses using such variables, use the chi-square test.

The chi-square test is used in two similar but distinct circumstances:

* Goodness-of-fit test to estimate how closely an observed distribution matches an expected distribution
* For estimating whether two random variables are independent.
* The Goodness-of-Fit Test

One of the more interesting goodness-of-fit applications of the chi-square test is to examine issues of fairness and cheating in games of chance, such as cards, dice, and roulette. Since such games usually involve wagering (betting), there is significant incentive for people to try to rig the games and allegations of missing cards, "loaded" dice, and "sticky" roulette wheels are all too common.

* Testing Independence

The other primary use of the chi-square test is to examine whether two variables are independent or not. What does it mean to be independent, in this sense? It means that the two factors are not related. Typically in social science research, we're interested in finding factors that are related - education and income, occupation and prestige, age and entrepreneurship. In this case, the chi-square can be used to assess whether two variables are independent or not.

More generally, we say that variable Y is "not correlated with" or "independent of" the variable X if more of one is not associated with more of another. If two categorical variables are correlated their values tend to move together, either in the same direction or in the opposite.

* **Example**

Examine statistically whether women paid their loans better than men.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Paid** | **Do not pay** | **Total** |
| **Women** | 46 | 71 | 117 |
| **Men** | 37 | 83 | 120 |
| **Total** | **83** | **154** | **237** |

**Step 1: Establish Hypotheses**

The first step of the chi-square test for independence is to establish hypotheses. The null hypothesis is that the two variables are independent - or, in this particular case that the likelihood of paying is the same for men and women. The alternative hypothesis to be tested is that the likelihood of paying the loan is not the same for men and women.

**Note:** The chi-square test only tests whether two variables are independent. It cannot test for significant differences (i.e. is greater or less). Therefore, using the chi-square test, we cannot evaluate directly the hypothesis that women paid their loans better than women; rather, the test (strictly speaking) can only test whether the two variables are independent or not.

**Step 2: Calculate the expected value for each cell of the table**

The key idea of the chi-square test for independence is a comparison of observed and expected values. How many of something was expected and how many was observed in the process? In the case of tabular data, however, we usually do not know what the distribution should look like. Expected values are therefore calculated based on the row and column totals from the table.

The expected value for each cell of the table can be calculated using the following formula:

(Row total x Column total) ÷ (Total for the table)

Using the formula, we can reconstruct our table as the following:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Paid** | **Paid (Expected value)** | **Do not pay** | **Do not pay (Expected value)** | **Total** |
| **Women** | 46 | 40.97 | 71 | 76.02 | 117 |
| **Men** | 37 | 42.03 | 83 | 77.97 | 120 |
| **Total** | **83** |  | **154** |  | **237** |

**Step 3: Calculate Chi-square statistic**

Chi-square statistic formula:

With these sets of figures, we calculate the chi-square statistic as follows:

+ + + = 1.87

**Step 4: Assess significance level**

To determine the significance level we need to know the "degrees of freedom." In the case of the chi-square test of independence, the number of degrees of freedom is equal to the number of columns in the table minus one multiplied by the number of rows in the table minus one. In the table above, the number of degrees of freedom = 1. We then compare the value calculated in the formula above to a standard set of tables. The value returned from the table is p < 20%. Thus, we cannot reject the null hypothesis. We therefore conclude that women are not significantly more likely to pay loans given them than men.

**Note:** P-values tell us the probability of our observations being due to chance (null hypothesis). The smaller the p-value (probability value), the stronger the evidence is in favour of the alternative hypothesis. P-values are usually determined using pre-determined significant levels under 0.05% or 0.01%. This means that under 0.05%, we are sure of our results being 95% reliable and under 0.01%, we are sure of our results being 99% reliable.

The T-Test

Introduction

How are results different for different groups? It's a question of central concern to researchers, because results vary and depend on a number of characteristics like where we live, how far we went in school, what kind of job we have. One goal of research is to accurately measure phenomena and document the levels of different features of society. However, we often want to use these measurements in evaluating specific hypotheses about these differences.

The **t-test** is a statistical tool to evaluate hypotheses about group-level differences in outcomes. Two different applications of the t-test in evaluating two kinds of hypotheses are:

1. The one-sample t-test, which compares the level of outcome for a group to a known standard.
2. The two-sample t-test, where the outcome levels of two groups are compared to each other.

History and Main Ideas

The t-test was developed by W. S. Gossett, a statistician employed at the Guinness brewery. However, because the brewery did not allow employees to publish their research, Gossett's work on the t-test appears under the name "Student" (so the t-test is sometimes called "Student's t-test"). Gossett was a chemist and was responsible for developing procedures for ensuring the similarity of batches of Guinness. The t-test was developed as a way of measuring how closely the yeast content of a particular batch of beer corresponded to the brewery's standard. One of the advantages of the t-test is that it can be applied to a relatively small number of cases. It was specifically designed to evaluate statistical differences for samples of 30 or less.

But the t-test is applied in other areas like the social world and agriculture. For example: how different are the WASSCE grades of 1st year Diploma students of Ohawu Agricultural College from the average WASSCE grades in 2014 in Ghana?

The same statistical methodology that compares a particular batch of beer to a standard can be used to compare how different any two batches are from each other (two samples). The test can be used to compare the yeast content of two kegs of beer brewed at separate times.

One-Sample T-Test

To reiterate, the one-sample t-test compares the mean score of a sample to a known value, usually the population mean (the average for the outcome of some population of interest). The basic idea of the test is a comparison of the average of the sample (observed average) and the population (expected average), with an adjustment for the number of cases in the sample and the standard deviation of the average.

* Example 1 – Improving yield of pepper

*One of the best indicators of a strong economy is its foreign exchange earnings. MOFA reports that although pepper production is only 5000ha or less, it accounts for nearly 2/3 of our agricultural foreign exchange earnings. In Ghana, farmers who live in Ohawu generally have lower yields of pepper per acre than those who do not live in Ohawu. While the average yield of pepper per acre in Ghana is approximately 3300kg, the average yield in Ohawu is 2800kg. Meanwhile Ohawu accounts for 30% of pepper production in Ghana. Recently, Ohawu Agricultural College introduced an innovative program to increase the yield of pepper in Ohawu. In the first year, 25 farmers in Ohawu participated in the program. Data drawn from the program reveals that weight of pepper per acre was 3075kg, with a standard deviation of 500kg.*

Has this program been effective at improving the yield of pepper in Ohawu?

* + - Step 1: Establish Hypotheses

**Null hypothesis** - the difference between pepper yield of farmers who participated in the program and pepper farmers in Ohawu is zero (0). Or the difference between the observed mean of pepper yield for program beneficiaries and the expected mean of yield for pepper farmers in Ohawu is zero.

**Alternative hypothesis** - the difference between the observed mean of pepper yield for program participants and the expected mean of yield for pepper farmers in Ohawu is not zero.

* + - Step 2: Calculate Test Statistic

Calculation of the test statistic requires four components:

1. The average of the sample (observed average)
2. The population average or other known value (expected average)
3. The standard deviation (SD) of the sample average
4. The number of observations.

With this example, the components are as follows:

1. Sample average = 3075kg
2. Population average of Ohawu pepper = 2800kg
3. Standard deviation (SD) of the sample average = 300kg
4. Number of observations = 25

With these four pieces of information, we calculate the t-statistic:

= = 0.898

* + - Step 3: Use the value of *t* to determine p-value

Compare the t-value you calculated with a standard table of t-values to determine whether the t-statistic reaches the threshold of statistical significance. Using the values of t = 0.898 and n (number of cases = 25) yields a p-value of 0.378. Generally speaking, we require p-values < 0.05 in order to reject the null hypothesis. With a value of 0.378, we cannot reject the null. In other words, the probability that our results are due to chance (or error) is 37.8%. That is very high and therefore, we conclude that the program did not successfully improve pepper yield. Meaning we used a confidence limit of 95%.

Two-Sample T-Test

We often want to know whether the means of two populations concerning some phenomena differ. For example, we want to compare two categories of some categorical variable (e.g., compare males and females) or two populations receiving different treatments in context of an experiment. The two-sample t-test is a hypothesis test for answering questions about the mean where the data are collected from two random samples of independent observations, each from an underlying normal distribution. The steps of conducting a two-sample t-test are quite similar to those of the one-sample test.

* Example 2- Treatment and control groups

In this example, rather than comparing the yield of a crop to the average of a whole, we will examine an intervention’s effect by comparing the yield of program participants with the yield of a group that did not (control).

To evaluate the effects of an intervention, program, or treatment, a group of subjects is divided into two groups. The group receiving the treatment to be evaluated is referred to as the treatment group, while those who do not are referred to as the control or comparison group. In this example, farmers who are part of the experimental program to increase yield are the treatment group. Define the control group.

* + - Step 1: Establish Hypotheses

Null hypothesis - the difference between the two groups is 0. Or the difference between the mean of the treatment group yield and the mean of the control group yield is zero.

Alternative hypothesis - the difference between the observed mean of treatment yield and the expected mean of non-intervention group (control) is not zero; therefore they are statistically different.

* + - Step 2: Calculate Test Statistic

1. The average of both sample (observed averages). Represented as: 1 and 2.
2. The standard deviation (SD) of both averages. Represented as: SD1 and SD2.
3. The number of observations in both populations, represented as: *n1* and *n2*.

From hospital records, we obtain the following values for these components:

|  |  |  |
| --- | --- | --- |
|  | **Treatment** | **Control** |
| **Average weight** | 3100kg | 2750kg |
| **SD** | 420 | 425 |
| **n** | 75 | 75 |

With these pieces of information, we calculate the t-statistic:

= = 5.07

* + - Step 3: use this value to determine p-value

Compare the t-value with a standard table of t-values to determine whether the t-statistic reaches the threshold of statistical significance. After checking from the standard table of t-values, the p-value is determined as 0.001, a score that forms our basis to reject the null hypothesis and conclude that the experiment made a difference.

ANALYSIS OF VARIANCE (ANOVA)

**The T-test compares variance between two groups using the t statistic but ANOVA uses the F statistic to compare variance between more than two groups using independent and dependent variables; under the assumption that treatment variances are the same and normally distributed. If the observed differences are a lot bigger than what is expected by chance, then there is statistical significance.**

**One-Way ANOVA:**ANOVA hypothesis test of 2 or more conditions (using population means) based on one independent variable (characteristic or factor) to determine if there is a difference between their means on a continuous dependent variable. One-way ANOVA cannot tell which specific groups were significantly different from each other; it only shows that at least two groups were different. Post-hoc tests, however, determine where differences lie between the specific groups.

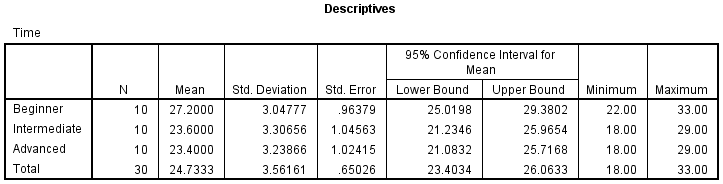
### Assumptions for One-Way ANOVA to be valid

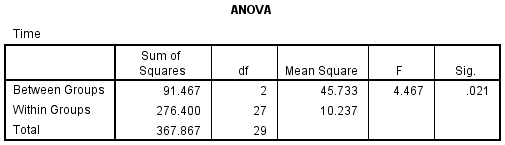
1. The dependent variable should be measured at the interval or ratio level (as continuous).
2. The independent variable should consist of two or more categorical, independent groups.
3. There should be independence of observations with random selection of samples from the population, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. If the data fails this assumption, use another statistical test instead of the one-way ANOVA (e.g., a repeated measures design).
4. There should be no significant outliers; they can have a negative effect on the one-way ANOVA, reducing the validity of your results. Outliers are simply single data points within your data that do not follow the usual pattern (e.g., in a study of 100 maize crop yields, where the mean weight was 108kg with only a small variation between crops, one crop with a weight of 236kg is very unusual).
5. The dependent variable should be approximately normally distributed for each category of the independent variable. We talk about the one-way ANOVA only requiring approximately normal data because it is quite "robust" to violations of normality, meaning that assumption can be a little violated and still provide valid results. Test for normality using the Shapiro-Wilk test of normality or Skewness and Kurtosis or Histograms or Normal Q-Q Plots.
6. The standard deviations (SD) of the populations for all groups are equal - this is sometimes referred to as an assumption of the homogeneity of variance. Again, we can represent this assumption for groups 1 through *n* as: = = = ... = . Check for homogeneity of variances using Levene's test for homogeneity of variances. If your data fails this assumption, carry out a Welch ANOVA instead of a one-way ANOVA, and a Games-Howell test instead of a Tukey post-hoc test.

**Two-Way ANOVA:** ANOVA hypothesis test that considers comparisons between populations based on 2 characteristics (variables/factors). Each characteristic can have multiple conditions.

**Note:** conditions = levels = treatments = groups but Independent variables = factors

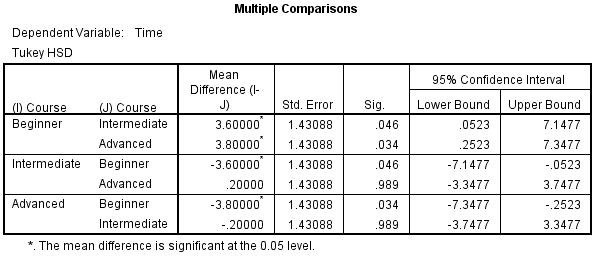
### Activity





**Note: When the p-value is less than the Alpha (α) level (margin of error), then there is significant difference between treatments. Generally speaking, we require p-values < 0.05 in order to reject the null hypothesis.**

**If Fcalc is greater than Fcrit then the Ho is rejected, and the means of one or more groups are different.   
If Fcalc is less than Fcrit then Ho stands; meaning the groups are not significantly different.**



There was a statistically significant difference between groups as determined by one-way ANOVA (F(2,27) = 4.467, p = 0.021). A Tukey post-hoc test revealed that the time to complete the problem was statistically significantly lower after taking the intermediate (23.6 ± 3.3 min, p = 0.046) and advanced (23.4 ± 3.2 min, p = 0.034) course compared to the beginners course (27.2 ± 3.0 min). There were no statistically significant differences between the intermediate and advanced groups (p = 0.989).

### ****Assignment****

**What is the difference between the z-score, p-value, t-statistic and F statistic? Under what conditions should they be used? How do you interpret them?**

### Quiz 1

1. Give 3 examples each of One-Way and Two-Way ANOVA experiments.
2. What conditions must exist in order to use ANOVA instead of Student’s T-test?
3. Under what circumstance do you use One-Way or Two-Way ANOVA?
4. What are the symbols for variance and standard deviation?

NOTE

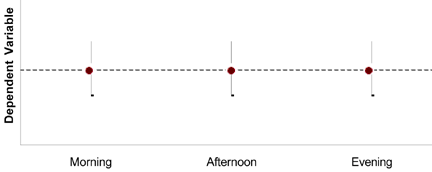
The t-test is a useful tool for comparing the means of two groups; with three or more groups, the t-test is not an effective statistical tool. ANOVA is the preferred statistical too to examine differences among the means of several different groups at once.

More generally, ANOVA is a statistical technique for assessing how nominal independent variables influence a continuous dependent variable under the assumption that treatment variances are the same and normally distributed.

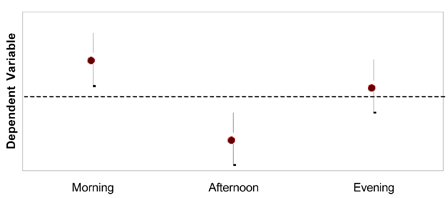
One-Way ANOVA Example

An experiment at OAC was conducted to identify the best time of watering cabbage: one in the morning, another in the afternoon, and a third in the evening. These treatments are carried out using the same litre of water per plot applied in the same way on same-sized plots with similar nutrient levels planted with the same type of seed from the same source.

The researcher is interested in whether there's any variation in how well the treatments performed. This is to find out whether busy city workers can integrate backyard cabbage production into their schedule which they can water at any time in the day.

Null hypothesis: the three treatments will have the same mean. In formula terms, if we use the symbol **μ** to represent the mean, the null hypothesis is expressed using the notation:

H0: **μ** morning = **μ** afternoon = **μ** evening

The graph above shows that all three groups have the same mean (all three points are on the dashed line) and all three groups have the same SD (noted by the fact that the line around the mean point for each group is the same size).

The alternate hypothesis is that all means will not be the same. It is important to note that it does not imply that all means will have different values. It is possible that some of the means could be the same, yet if they are not all identical, we would reject the null hypothesis. Therefore, the alternative hypothesis is that not all means are equal. This is shown graphically in the graph above.

ANOVA allows us to separate the total variability in the outcome into two parts: variability within groups and variability between groups. We can calculate an ANOVA statistic called F statistic to evaluate the hypothesis. Then compare this calculated F value to a standard table with values for the F distribution to determine the significance level for the F value.

EXPERIMENTAL DESIGN

Experimental designs arose from the way in which experimental units are grouped or classified.

**Definition:** An experimental design is the actual layout of an experiment in respect of its number of replication and plots and their special relation to one another, irrespective of whether the allocation of treatment is made at random or otherwise.

Types of experimental design

The principal difference among experimental designs is the way in which plots are grouped or classified by treatments, blocks, rows, main plots and the like. Some of the main experimental designs are:

1. The Completely Randomized Design
2. The Randomized Complete Block Design
3. The Latin Square Design
4. The Graeco Latin Square design
5. The Split Plot Design

Out of these, the 2nd, 3rd and 5th designs are commonly.

* The Completely Randomized Design

This design is the simplest and is set up by assigning treatments at random to a determined set of plots. Usually this is not the most efficient design for field trials with plants. However, it may be the most workable arrangement for testing certain types of treatments on animals. Any number of treatments may be tested in this design.

* + - Layout of plots:

1. A number of plots are marked out on a field. (This design has no blocks).
2. The number of experimental units laid out is equal to the number of treatments times the desired number of replications.
3. Individual plots are usually rectangular areas of identical shape.
4. The treatments are randomly allocated to the plots such that each treatment occurs in the design according to the number of replications.

**Disadvantage:** The design is incapable of estimating the experimental error with a high degree of precision.

**Advantages:**

1. It is simple and flexible in terms of randomisation of treatments and analysis of data.
2. It maximises the degree of freedom for estimating experimental error
3. It minimises the F-value required for statistical analysis.
4. Loss of information as a result of missing data is relatively small compared with other designs.
   * + Analysis of variance (ANOVA)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Plot Number** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **Variety** | B | A | A | B | A | B | B | A | A | B |
| **Yield** | 12 | 10 | 14 | 18 | 15 | 16 | 14 | 20 | 16 | 11 |

To get the results analyzed, the plot yields should be grouped by varieties/treatments as shown below

|  |  |  |  |
| --- | --- | --- | --- |
| **Variety/Treatment** | **Yield per plot (kg)** | **Total ()** | **Mean ()** |
| **A** | 10 14 15 20 16 | 75 (TA) | 15 |
| **B** | 12 18 16 14 11 | 71 (TB) | 14. 2 |
| **Grand Total** | | **146** | **14. 6** |
| **Grand Mean** | | |

**ANOVA tabulation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source of variance** | **Degree of freedom** | **Sum of squares** | **Mean square** | **Observed F/**  **Computed F** | **Required F**  **5% 1%** | |
| Total | 9 | 86. 4 |  |  |  |  |
| Treatments | 1 | 1.6 | 1. 6 | 0. 2 ***ns*** | 5.32 | 11. 26 |
| Error | 8 | 84.8 | 10.6 |  |  |  |

**Degree of freedom**

1. Total = Number of experimental units – 1 = 10 – 1 = 9
2. Treatments = Number of treatments – 1 = 2 – 1 = 1
3. Error = Number of treatments(number of replications – 1) = 2(5-1) = 8

**Correction factor**

CF = () 2 =T2 = (146)2

n n 10 = 2131.6

Where n = total numbers of experimental plots, and T = sum of the yield from all plots.

**Sum of squares**

1. Total = SSTotal = ∑ 2 - ( ∑x) 2  = ∑ 2  - CF = ((10)2 + (14)2 + (15)2 + ... + (11)2) - CF

n

= 2218 – 2131. 6 **= 86.4**

Where n = total numbers of experimental plots

1. Treatments = SSTreatments = (TA) 2 + (TB) 2 – CF = (TA) 2 + (TB) 2 – CF **=** (75)2 + (71)2 – CF = 2133. 2 – 2131. 6

n n n 5

= **1. 6**

**Note:** Where n = number of replications (plots) per treatment and TA /TB = total yield per treatment.

1. Error = SSError = SSTotal – SSTreatments = 86. 4 – 1. 6 = **84. 8**

Now calculate the mean square (MS) for each source of variation by dividing each SS by its corresponding degree of freedom (d. f.).

**Mean squares**

1. MSTreatments = SSTreatments ÷ (t- 1) Where t-1 = Degree of freedom for treatments =1.6 ÷ 1 = **1. 6**
2. MSError = SSError ÷ t(r–1) Where t= no. of treatments and r= no. of replications per treatment

= 84. 8 ÷ 2(5 – 1) = **10. 6**

**F values**

1. Observed F-value = MSTreatments ÷ MSError = 1.6 ÷ 10.6 = **0. 2**

**Note:** Observed F or Computed F is also called Calculated F.

1. Required F-value: To obtain the required F-value, refer to tables of degree of freedom Use d.f. treatment on the Horizontal and d. f. error on the Vertical. Thus for our example, we have Horizontal = 1, and vertical = 8. The values are 5.32 for the 5% level of significance and 11.26 for the 1% level.

**Note:** Required F is also called Tabulated F or F Critical.

The F values are for testing statistical significance of treatment differences. If the Observed F-value is less than the Required F-value at the chosen significance level (e.g. 5% or 1%), then there is no statistically significant difference between treatments. Such a result is indicated by placing***ns*** on the computed F-value in the analysis of variance. However, when the Observed F-value is greater than the Required F-value at the chosen significance level, then there is statistically significant difference between treatments. Such a result is indicated by placing***s*** on the computed F-value in the analysis of variance.

For our example, the computed F-value of 0.2 is smaller than the tabulated F-value at the 5% level of significance, the treatment difference is therefore said to be *non significant*.

In a situation where the F value proved to be significant, then the Coefficient of Variance could be calculated as follows: CV = (MSError ÷ Grand mean) x 100 = (10.6 ÷ 14.6) x 100 = **22.6%**

The CV indicates the degree of precision with which the treatments are compared and is a good index of the reliability of the experiment. It expresses the experimental error as a percentage of the mean; thus the higher the CV value, the lower the reliability of the experiment. The CV value is generally placed below the analysis of variance (ANOVA) table.

THE RANDOMIZED COMPLETE BLOCK DESIGN

Sometimes called the Randomized Block Design, it is set up by assigning treatments at random to a group of experimental units called the block or replication. Block is the preferred term since it is better differentiated from replication in the design with no blocks (the Completely Randomized Design). It keeps variability among plots as small as possible but maximises differences among blocks. Blocks are as compact as possible to reduce to a great minimum, soil and other environmental differences.

* + - Layout of plots

1. The site for the experiment is divided into a number of blocks and plots of equal dimensions**.** The relative position of the block is immaterial. Soil uniformity within any one block is the primary consideration.
2. The number of blocks should be the same as the number of replications.
3. The number of plots in a block should be equal to the number of treatments.
4. Each treatment must appear once in a block.
5. Individual plots are usually rectangular areas of identical shape and size.
6. The number of replications should not be less than four (4). More accurate results are likely to be obtained with a greater number of replications.
7. Treatments are assigned to plots within each block in a purely random manner. That is, a separate randomisation is made for each block.
8. Treatments can be as many as possible.
9. The number of plots laid out = the number of treatments x the number of replications per treatment.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **I** | **II** | **III** | **IV** | **V** | | D | A | C | C | A | | A | D | D | B | C | | B | C | B | D | D | | C | B | A | A | B | | Low fertility High fertility  **Four treatment replicated five times in a Randomized Complete Block Design** | | | | | |

ANALYSIS OF VARIANCE FOR RANDOMISED COMPLETE BLOCK DESIGN

Yields grouped by treatments and blocks

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **TREATMENT** | **B L O C K**  I II III IV | | | | **TREATMENT TOTAL** | **TREATMENT**  **MEAN** | |
| **A** | 1. 0 | 0. 4 | 1. 2 | 0. 8 | 3. 4 | | 0. 85 |
| **B** | 3. 0 | 1. 2 | 3. 6 | 2. 6 | 10. 4 | | 2. 60 |
| **C** | 3. 0 | 1. 4 | 4. 0 | 3. 0 | 11. 4 | | 2. 85 |
| **D** | 2. 4 | 1. 0 | 2. 4 | 1. 6 | 7. 4 | | 1. 85 |
| **TOTAL** | 9. 4 | 4. 0 | 11. 2 | 8. 0 | 32. 6 | | 8. 15 |

**Correction Factor** (CF) = 2 ÷ N = 32. 62 ÷ 16 = 1062.76 ÷16 = **66.42**

**Sum of squares**

**SSTotal** = 2) - CF

2) = (1.02+0.42 +1.22 + 0.82) + (3. 02 + 1.22 + 3. 62 + 2. 62) +

(3. 02 + 1.42 + 4.02 + 3.02) + (2. 42 + 1.02 + 2.42 + 1.62)

= 3.24 + 30. 16 + 35.96 +15.08 = 84.44 🡺 **SSTotal** = 84.44 – 66.42 = **18.02**

**SSBlocks** = (2) ÷ n) - CF = ((9. 42 + 4. 02 + 11. 22 + 8. 02) ÷ 4) – 66.42 = 73.45 – 66.42 = **7.03**

**SSTreatments** = (2) ÷ n) – CF = ((3. 42 + 10. 42 + 11. 42 + 7. 42) ÷ 4) – 66.42 = 76. 11 – 66. 42 = **9.69**

**SSError** = SSTotal – SS­Blocks – SSTreatments = 18.02 – 7.03 – 9.69 = **1.3**

**Degrees of freedom**

Total d.f. = (No. of blocks x no. of treatments) – 1 = 16-1 = 15

Blocks d.f. = No. of blocks – 1 = 4-1 = 3

Treatments d.f. = No. of treatments – 1 = 4-1 = 3

Error d.f. = Treatments d.f. x Blocks d.f. = 3 x 3 = 9

**A N O V A T A B L E**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variance | Degree of freedom | Sum of squares | Mean square | Observed F | Required F  5% 1% | |
| Total | 15 | 18. 02 |  | 22.36 | 3. 86 | 6. 99\*\* |
| Blocks | 3 | 7. 03 | 2. 343 |
| Treatments | 3 | 9. 69 | 3. 23 |
| Error | 9 | 1. 3 | 0. 1444 |

Mean squares = Sum of squares ÷ the degree of freedom

MSBlocks = 7.03 ÷3 = **2. 343** MSTreatments = 9.69 ÷ 3 = 3.23

MSError = 1.3 ÷ 9 = 0.14444 Observed F-value = MSTreatments ÷ MSError = 3.23 ÷ 0.144444 = 22.36

**Required F Values:** Read from tables. Thus readings for degree of freedom for treatment being 3 and for error being 9 are 3.86 at 5% and 6. 99 at 1%.

An indication that the experiment is highly significant since the observed F value is greater than the required value at both 5% and 1% level. The experiment would have been insignificant if the F- value required is greater than the observed.

THE LATIN SQUARE DESIGN

Certain fields chosen for experiments may exhibit variation in fertility trend. Such variation could be eliminated using the Latin Square experimental arrangement. In this design, the randomisations of treatments are restricted further by grouping them into rows as well as columns. The design simultaneously controls fertility variations in two directions of replications at right angles.

* + - Layout of plots

1. In this design, the number of replications = the number of treatments.
2. The experimental area is divided into units or plots arranged in a square in such a manner that there are as many plots in each row as there are in columns, this number = the number of treatments.
3. The treatments are randomly assigned to plots such that every treatment occurs once in each row and once in each column.
4. The Latin Square gives more precision when the plots (or units) have equal (square) dimensions or almost a square as opposed to long narrow strips.
5. The Latin Square requires at least as many replications as there are treatments and therefore is not practicable for experiments with large number of treatments. The most commonly used are those having four (4) to eight (8) treatments.
6. When the number of treatments is few, < 4, the Latin Square is not efficient for measuring small differences. For small number of treatments, however, two or more Latin Squares are laid out to secure adequate replication.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **COLUMNS** | | | | | | |
| **ROWS** |  | **I** | **II** | **III** | **IV** | **V** | **VI** |
| **I** | B | D | E | F | A | C |
| **II** | C | E | A | D | F | B |
| **III** | A | F | C | B | E | D |
| **IV** | D | A | F | C | B | E |
| **V** | F | B | D | E | C | A |
| **VI** | E | C | B | A | D | F |

**A six by six Latin Square Design: Letters indicate treatments.**

**Advantages:**

1. Treatments are replicated in two different ways – rows and columns, to increase precision.
2. Soil differences are removed completely in a Latin square.

**Disadvantages:**

1. The number of replications is limited by the number of treatments being studied.
2. With few treatments (for example, 3 or 4), the Latin Square is less efficient in removing experimental error as with the Randomized Complete Block with more replications.
3. It is less flexible in character than the Randomized Complete Block

LATIN SQUARE DESIGN ANALYSIS

**Yields of five (5) treatments of Cowpea in kg/plot**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ROWS** | **COLUMNS** | | | | | |
|  | **I** | **II** | **III** | **IV** | **V** |
| **I** | A | B | C | D | E |
| **2** | **5** | **3** | **1** | **6** |
| **II** | B | C | D | E | A |
| **5** | **3** | **1** | **5** | **2** |
| **III** | C | D | E | A | B |
| **4** | **2** | **6** | **3** | **4** |
| **IV** | D | E | A | B | C |
| **1** | **6** | **3** | **4** | **3** |
| **V** | E | A | B | C | D |
| **5** | **3** | **5** | **3** | **2** |

**Yields Grouped by Row and Column**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Column** | | | | | **Row Total**  **(T­row)** |
| **I** | **II** | **III** | **IV** | **V** |
| **I** | 2. 0 | 5. 0 | 3. 0 | 1. 0 | 6. 0 | 17. 0 |
| **II** | 5. 0 | 3. 0 | 1. 0 | 5. 0 | 2. 0 | 16. 0 |
| **III** | 4. 0 | 2. 0 | 6. 0 | 3. 0 | 4. 0 | 19. 0 |
| **IV** | 1. 0 | 6. 0 | 3. 0 | 4. 0 | 3. 0 | 17. 0 |
| **V** | 5. 0 | 3. 0 | 5. 0 | 3. 0 | 2. 0 | 18. 0 |
| **Column Total**  **(Tcolumn)** | **17. 0** | **19. 0** | **18. 0** | **16. 0** | **17. 0** | **87. 0** |

**Yields Grouped by Treatment and Row**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Treatment** | **Row** | | | | | **Treatment Total (Tt)** | **Treatment Mean** |
| **I** | **II** | **III** | **IV** | **V** |
| **A** | 2. 0 | 2. 0 | 3. 0 | 3. 0 | 3. 0 | **13** | 2. 6 |
| **B** | 5. 0 | 5. 0 | 4. 0 | 4. 0 | 5. 0 | **23** | 4. 6 |
| **C** | 3. 0 | 3. 0 | 4. 0 | 3. 0 | 3. 0 | **16** | 3. 2 |
| **D** | 1. 0 | 1. 0 | 2. 0 | 1. 0 | 2. 0 | **7** | 1. 4 |
| **E** | 6. 0 | 5. 0 | 6. 0 | 6. 0 | 5. 0 | **28** | 5. 6 |

**A N O V A T A B L E**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source of Variance** | **d. f** | **S S** | **M S** | **Observed F** | **Required F** | |
| **Total** | 24 | 60. 24 |  | 47.1 | **5%** | **1%** |
| **Rows** | 4 | 1. 04 | 0. 26 | 3. 26 | 5. 41 |
| **Columns** | 4 | 1. 04 | 0. 26 |
| **Treatments** | 4 | 54. 64 | 13. 66 |
| **Error** | 12 | 3. 52 | 0. 29 |

**Degrees of freedom** Total = Total no. of treatments – 1

Rows = Total no. of rows – 1. Columns = Total no. of columns – 1

Treatments = Total no. of treatments – 1 Error = (no. of treatments -1)(no. of treatments - 2)

**Correction Factor** C.F. = (87)2 ÷ 25 = 302.76

**Sum of squares**

SSTotal = ∑2) – C.F. = 2. 02 + 5. 02 +……………..… + (2. 0)2 – CF = 363 -302.76 = 60. 24

SSRows = (∑(Tr2) ÷ n) – C.F. = ((172 + 162 + 192 + 172 + 182) ÷ 5) – 302.76 = 1. 04

SSColumns = (∑(Tc2) ÷ n) – C.F. = ((172 + 192 + 182 + 162 + 172) ÷ 5) - 302.76 = 1. 04

SSTreatments = (∑(Tt2) ÷ n) – C.F. = 132 + 232 + 162 + 72 + 282 – CF = 54. 64

SError = SSTotal – SSRows – SSColumns – SSTreatments =60. 24 – (1. 04 + 1. 04 + 54. 64) = 3. 52

**Mean squares**

MSRows = SSRows ÷ Rows degree of freedom = 1.04 ÷ 4 = 0.26

MSColumns = SSColumns ÷ Columns degree of freedom = 1.04 ÷ 4 = 0.26

MSTreatments = SSTreatments ÷ Treatments degree of freedom = 54.64 ÷ 4 = 13.66

MSError = SS­Error ÷ Error degree of freedom = 3.52 ÷ 12 = 0.26

RANDOMISED COMPLETE BLOCK VERSUS LATIN SQUARE

The Randomized Complete Block Design is superior to the Latin Square in many ways.

1. A wide range of treatments can be tested in a Randomised Complete Block since there is no restriction on the number of replications as there are treatments and is thus suitable for more than 8 treatments.
2. The Latin Square is not suitable with too few or too many treatments. With few treatments, the number of replications is found inadequate. With many treatments, the number of replications is unduly increased to make the design practically unworkable.
3. In a Randomised Complete Block, if there is an attack of some pest or disease on one or two of the blocks, the data for these blocks can easily be omitted without any complications in the analysis; a more complicated analysis is necessary under similar circumstances in a Latin Square.
4. In the field, the Randomised Complete Block is easier to manage than the Latin Square. It can be accommodated equally well in a field of any other shape from square to rectangular; in a Latin Square the field must be approximately square or rectangular.
5. If there are any obstacles in the experimental site, the blocks can easily be separated in a Randomised Complete Block; this is not practicable in a Latin Square.
6. When there is simultaneous trend of fertility variation at right angles (i.e. diagonal fertility trend), the Latin Square is likely to be more efficient.
7. Statistical analysis of data (variance) is more flexible in the Randomised Complete Block.
8. The Latin Square arrangement is also better in some branches of agricultural research; for example, in animal research where two factors are being studied (e.g. litter and body weight differences may be controlled in trials with guinea pigs by assigning them to rows and columns of a Latin Square).

* Factorial experiment

A factorial experiment is one carried out to test two or more factors simultaneously. It is also an experiment in which the treatments consist of all possible combinations of the selected levels in two or more factors. For example, an experiment involving two factors, each at two levels such as two broad leaf weeds and two doses of herbicide, is referred to as a 2 x 2 or 22 factorial experiment.

* Single factor experiment

Experiments in which a single factor varies while all others are kept constant are single-factor experiments. Treatments in such experiments consist solely of the different levels of one variable factor. All other factors are applied uniformly to all plots at a single prescribed level.

Examples of single factor experiments:

1. Varietal trial where several varieties of maize are planted and all management factors such as fertilizer, insect control, and water management, are applied uniformly.
2. Fertilizer trials where several rates of a single fertilizer element are tested.
3. Insecticide trials where several insecticides are tested.
4. Plant population trials where several plant densities are tested.